remote sensing

MDPI

Article

Inferring Grassland Drought Stress with Unsupervised Learning from Airborne Hyperspectral VNIR Imagery

Floris Hermanns [1,*], Felix Pohl [2], Corinna Rebmann [2], Gundula Schulz [3], Ulrike Werban [4] and Angela Lausch [1,5]

1 Helmholtz Center for Environmental Research—UFZ, Department of Computational Landscape Ecology, Permoserstr. 15, 04318 Leipzig, Germany; angela.lausch@ufz.de
2 Helmholtz Center for Environmental Research—UFZ, Department of Computational Hydrosystems, Permoserstr. 15, 04318 Leipzig, Germany; felix.pohl@ufz.de (F.P.); corinna.rebmann@ufz.de (C.R.)
3 Helmholtz Center for Environmental Research—UFZ, Department of Remote Sensing, Permoserstr. 15, 04318 Leipzig, Germany; gundula.schulz@ufz.de
4 Helmholtz Center for Environmental Research—UFZ, Department of Monitoring and Exploration Technologies, Permoserstr. 15, 04318 Leipzig, Germany; ulrike.werban@ufz.de
5 Lab for Landscape Ecology, Department of Geography, Humboldt University of Berlin, Rudower Chaussee 16, 12489 Berlin, Germany
* Correspondence: floris.hermanns@ufz.de

Abstract: The 2018–2019 Central European drought had a grave impact on natural and managed ecosystems, affecting their health and productivity. We examined patterns in hyperspectral VNIR imagery using an unsupervised learning approach to improve ecosystem monitoring and the understanding of grassland drought responses. The main objectives of this study were (1) to evaluate the application of simplex volume maximisation (SiVM), an unsupervised learning method, for the detection of grassland drought stress in high-dimensional remote sensing data at the ecosystem scale and (2) to analyse the contributions of different spectral plant and soil traits to the computed stress signal. The drought status of the research site was assessed with a non-parametric standardised precipitation–evapotranspiration index (SPEI) and soil moisture measurements. We used airborne HySpex VNIR-1800 data from spring 2018 and 2019 to compare vegetation condition at the onset of the drought with the state after one year. SiVM, an interpretable matrix factorisation technique, was used to derive typical extreme spectra (archetypes) from the hyperspectral data. The classification of archetypes allowed for the inference of qualitative drought stress levels. The results were evaluated using a set of geophysical measurements and vegetation indices as proxy variables for drought-inhibited vegetation growth. The successful application of SiVM for grassland stress detection at the ecosystem canopy scale was verified in a correlation analysis. The predictor importance was assessed with boosted beta regression. In the resulting interannual stress model, carotenoid-related variables had among the highest coefficient values. The significance of the photochemical reflectance index that uses 512 nm as reference wavelength (PRI_{512}) demonstrates the value of combining imaging spectrometry and unsupervised learning for the monitoring of vegetation stress. It also shows the potential of archetypical reflectance spectra to be used for the remote estimation of photosynthetic efficiency. More conclusive results could be achieved by using vegetation measurements instead of proxy variables for evaluation. It must also be investigated how the method can be generalised across ecosystems.

Keywords: vegetation stress detection; unsupervised machine learning; drought; hyperspectral VNIR data; monitoring solutions; photochemical reflectance index; pattern recognition; plant traits; grassland; imaging spectrometry

Citation: Hermanns, F.; Pohl, F.; Rebmann, C.; Schulz, G.; Werban, U.; Lausch, A. Inferring Grassland Drought Stress with Unsupervised Learning from Airborne Hyperspectral VNIR Imagery. *Remote Sens.* **2021**, *13*, 1885. https://doi.org/10.3390/rs13101885

Academic Editor: Yoshio Inoue

Received: 30 March 2021
Accepted: 2 May 2021
Published: 11 May 2021

1. Introduction

The changing background climate state is projected to lead to an increase in the frequency and severity of extreme weather and climate events in many parts of the world [1–3].

In Europe, prolonged droughts are among the most damaging natural hazards, both socioeconomically and environmentally [4]. Severe droughts impair agri- and silvicultural productivity, and they can have longer-term effects on vegetation health [5]. Hydrometeorological extremes are also a major contributor to interannual carbon uptake variability [6], adding to the instability of the strength of the terrestrial carbon sink. Pre-symptomatic vegetation stress monitoring is essential for analysing changes in plant functioning linked to extreme events like the 2018–2019 Central European drought on a broader scale. Remote sensing techniques for fast and reliable stress monitoring become increasingly valuable. Early vegetation stress detection is not only useful for improving land surface and ecosystem models, but also an important tool in a variety of fields, such as precision agriculture [7], forest resource management [8], ecosystem services maintenance [9], or biodiversity and geodiversity conservation [10,11].

At present, a wide range of remote sensing technologies is available for vegetation stress and productivity monitoring, which is attributable to the success of open research data and algorithms [12], among other things. Multi- and hyperspectral optical sensors collect data from the visible, infrared, and thermal wavelength domains. Active sensors, like synthetic-aperture radar, radiometers, and lidar, provide information from other regions of the electromagnetic spectrum [13] or assess vegetation structure [14]. Multi-angular remote sensing [15] and model-based approaches [16] enable the estimation of key variables, like sun-induced fluorescence [17]. The fusion of multisensoral datasets is increasingly used for tasks, such as correcting optical measurements that are influenced by canopy structure and viewing geometry [18], or increasing spatiotemporal resolution of imagery [19].

The narrow wavelengths that are typically recorded in the VNIR (400–1300 nm) and SWIR (1300–2500 nm) regions by imaging spectrometers hold information regarding pigment composition, structure, and water content of vegetation [20], which are influenced by plant responses to different stressors [21]. The identification of empirical relationships between pigment concentration and a combination of hyperspectral bands or derivative spectra enabled the development of spectral indices, thus spectral features of interest could be located [22]. When compared to broadband indices, like the NDVI, narrowband indices are more specific, but also more afflicted by factors, such as canopy structure, illumination, or atmospheric conditions [23]. Therefore, the established relationships cannot simply be extrapolated beyond the biological systems or conditions that they were developed for [24]. A notable exception is the photochemical reflectance index (PRI), which is a reliable proxy for photosynthetic efficiency over a range of species and biomes [25]. Full spectrum analysis methods are a way to circumvent the extrapolation limits of narrowband indices. By integrating the signals of stress responses from different spectral regions, such analyses reduce the uncertainty in predictions [22] and they achieve earlier stress detection, but require researchers to cope with a low signal-to-noise ratio [26].

Methods from the field of machine learning are increasingly used for evaluating the high-dimensional data cubes that are generated by hyperspectral sensors. Supervised learning algorithms of varying complexity can be used to fit statistical models with high predictive power when labeled observation data exists [7,27,28]. If no reference data is available, unsupervised learning algorithms can be used to infer latent variables, thus reducing the dimensionality of hyperspectral datasets in order to detect informative patterns [29]. Matrix factorisation algorithms are a popular group of techniques for latent variable inference. A new descriptive representation of a high-dimensional input matrix is generated by decomposing it into a matrix of latent components and a coefficient matrix [30]. Various matrix factorisation methods have been developed and used in all kinds of data mining applications [31–34]. They have also been brought to use for different remote sensing tasks, such as image fusion, unmixing, or feature classification [35–39].

Simplex volume maximisation (SiVM) is a matrix factorisation algorithm that combines the advantage of intuitively interpretable results with computational efficiency [40], which makes it a promising tool for monitoring tasks. Here, data are clustered by typical

extreme spectra (archetypes), which can be used for estimating vegetation stress levels. However, as far as we know, the method has only been used in small-scale experimental settings. Therefore, this study aims at (1) evaluating the application of SiVM to ecosystem scale hyperspectral VNIR data of a grassland site for qualitative drought stress classification and (2) analysing the contributions of different evaluation variables to the inferred stress signal.

Image acquisition was conducted during two aircraft campaigns. The first in spring 2018 at the onset of the 2018–2019 European drought, and the second in spring 2019 after one year of prolonged drought conditions. The drought status of the grassland research site was assessed with a non-parametric implementation of the standardised precipitation-evapotranspiration index (SPEI) [41] and soil moisture measurements. Because no vegetation measurements were made, statistical evaluation was based on a set of plant and soil trait proxy variables, including a number of established vegetation indices that are sensitive to drought-inhibited vegetation growth. After a correlation analysis, different models were fit for explanatory variable analysis using boosted beta regression.

2. Materials and Methods

2.1. Study Site and Drought Status

The intensive research site "Am Grossen Bruch" (DE-GsB) is part of the Bode catchment hydrological observatory that is a component of the German TERrestrial ENvironmental Observatories (TERENO) [42,43]. TERENO sites are equipped with numerous sensors to monitor ecosystem behaviour and status, such as carbon, water, and energy fluxes with the eddy covariance method, distributed soil moisture, as well as biotic and abiotic drivers. DE-GsB is also an associated site of the Integrated Carbon Observation System Research Infrastructure (ICOS) that aims to provide consistent long-term measurements of sources and sinks of greenhouse gases [44]. Being located in the Central German Lowland in the state of Saxony-Anhalt (see Figure 1), the region surrounding DE-GsB is characterised by a drier subcontinental climate, which is partly due to the rain shadow of the Harz mountains. The area has a negative climatic water balance [45], which makes it susceptible to more severe impacts from drought events. DE-GsB is a mesophilic grassland that is regularly used for cattle grazing. It is located near a water channel (52.03°N, 11.10°E) and is affected by seasonal flooding. Testing SiVM at the ecosystem scale with data from a flat, mixed grassland has certain advantages, such as minimizing the influence of uncontrolled variables, like species composition, canopy structure, and leaf angle distribution.

We assessed the hydrological conditions of DE-GsB with the SPEI, which is used for analysing anomalies in the climatic water balance from a distributional perspective. Thus, it can be used for the standardised detection of droughts and comparisons of drought severity in a statistically robust way [46]. We used a non-parametric approach to determine an empirical distribution function using a kernel density estimator, following the method that was described by Vergni et al. [47]. Gridded (1×1 km) daily meteorological data from a follow-up product of the High-Resolution Dataset of Water Fluxes and States for Germany [48] were used as input data, covering a time span from 1947–2019. These forcing data itself are derived from weather station data from the German Meteorological Service (Deutscher Wetterdienst, DWD), which were interpolated with external drift kriging, using terrain elevation as the drift variable. We used a Gaussian kernel for SPEI calculation and optimised the bandwidth selection with cross validation [49]. In addition, we evaluated continuous soil water content (SWC) data that were derived from a profile with TDR measurements in three depths (CS616, Campbell Scientific, Shepshed, UK) to assess the severity of the local soil water deficiency resulting from the prolonged drought. This was done to determine whether the two images showed contrasting environmental conditions, justifying factorisation with SiVM.

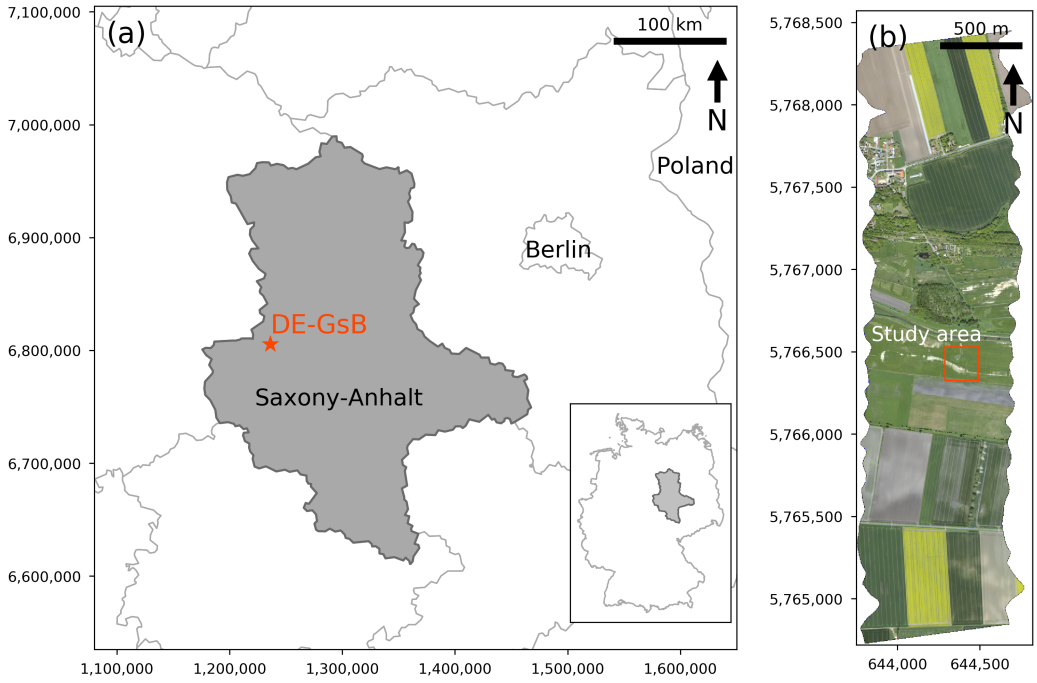


Figure 1. (**a**) Location of the research site "Am Grossen Bruch" (DE-GsB) within Germany (CRS: EPSG:3857). (**b**) RGB representation of the 2018 hyperspectral flight strip showing the surroundings of the study area (CRS: EPSG:32632).

2.2. Airborne Data

Flight campaigns for hyperspectral image acquisition on the DE-GsB site were conducted on 7 May 2018 and 23 April 2019. The images were collected using a HySpex VNIR sensor (NEO, Oslo, Norway) mounted on a Cessna 207 airplane from an altitude of 3 km a.g.l between 11:30 and 13:30 CEST under clear sky conditions. The spectra were recorded over the range 409–989 nm with a spectral resolution of 3.2 nm (see Table 1). Raw HySpex image blocks were radiometrically and geometrically corrected for surface reflectance with HySpexRAD and HySpexNAV (NEO). Georeferencing and calibration were performed with PARGE [50] using a digital elevation model (DEM) with an original resolution of 10 m obtained from the German Federal Agency for Cartography and Geodesy (Bundesamt für Kartographie und Geodäsie, BKG) and ground control points set in ENVI (Harris Geospatial Solutions, Boulder, CO, USA). Atmospheric correction was done with ATCOR4 [51] with visibility values from the nearby (10 km distance) Ummendorf DWD weather station [52]. The resulting image cubes had a spatial resolution of 0.5 m (2018) and 0.4 m (2019).

The atmosphere-corrected hyperspectral imagery went through the following preprocessing routines: The 2018 dataset was resampled to 0.4 m using cubic spline interpolation to align the spatial resolution and extent of both datasets. The 0.4 m target resolution was selected to increase the sample size for the statistical evaluation. The data were tested for noisy bands and bad pixels with a correlation-based approach [53]. The bands were deemed to be noisy when the Pearson correlation coefficient to adjacent bands/pixels fell below 0.8. Bad pixel candidates were selected as pixels with a value of at least 80% of the maximum band value. If the candidate value exceeded the maximum value of the same pixel in adjacent bands by 90% or more, the pixel was considered to be abnormal. Pixels with zero reflectance in all bands were also classified as bad pixels. All of the identified noisy bands were removed from all image cubes, resulting in the removal of the first seven of 182 bands from all data sets. Bad pixels were masked and excluded from factorisation. Except for the statistical evaluation of results, all of the data preprocessing and analyses were implemented in Python 3.7.3 (Python Software Foundation, https://www.python.org/, accessed on 23 February 2021).

Table 1. Specification of the hyperspectral sensor used for image acquisition.

Recording Date	Ground Resolution m	Field of View °	Swath m	Spectral Range nm	Spectral Resolution nm	Sensor	Platform
7 May 2018	0.5	16	508	409–989	3.2	HySpex VNIR-1800	Cessna 207
23 April 2019	0.4	16	508	409–989	3.2	HySpex VNIR-1800	Cessna 207

2.3. Geophysical Data

The geophysical measurements originate from a campaign in March 2014 that was conducted following a 10 day period of heat and no precipitation. Measurements were conducted with electromagnetic induction (EMI) sensors, which establish magnetic fields at the soil surface and measure the induced magnetic field of subsurface materials [54]. The sensor models EM38DD and EM31MK2 from Geonics, Canada were used (EM38 and EM31 in the following). Additionally, gamma ray (GR) emissions of Thorium ^{232}Th were measured using a portable gamma ray spectrometer from GF instruments. All of the sensors were attached to sledges and pulled over the research site with GPS positioning to measure apparent electrical conductivity (ECa) and nuclide concentrations [55].

ECa is controlled by a number of soil properties and states, such as soil texture, layering, or porosity [56], which can influence vegetation ecophysiology indirectly. Concerning soil moisture, Martini et al. [57] concluded that there is a complex interplay between factors controlling ECa and soil moisture, but not in any case a direct link. EM31 and EM38 measurements were taken in vertical coil orientation, with an effective penetration depth up to 1.5 m (EM38) and 4.0 m (EM31), respectively [55]. The gamma ray measurements are related to properties that are influenced by the source rock, soil genesis, management (e.g., pH), and soil moisture. For spatial interpolation, the transects were subsampled and then examined in a variogram analysis to find the best linear unbiased estimator for ordinary kriging using the GSTools module (v1.2.1) [58].

2.4. Unsupervised Classification with Simplex Volume Maximisation

Figure 2 provides an overview of the analysis procedure. SiVM allows the calculation of convexity constrained latent components [40], similar to precursor algorithms like Convex-NMF (C-NMF) [59] or Archetypal Analysis (AA) [60]. In the case of clustering of hyperspectral data, convex latent components are desirable as they are identical with actual measured data points, assigning them a physical meaning [61]. Furthermore, in contrast to common clustering methods, like k-means or DBSCAN, whose cluster centers represent average values of certain data regions, convexity constrained factorisation yields the characteristic extreme data points (archetypes) of a dataset [40]. These properties facilitate the interpretability of the resulting latent variable model. One major advantage of SiVM over similar factorisation algorithms is that it runs in linear time. This feature makes it a suitable algorithm for processing large amounts of data, e.g., hyperspectral image cubes at landscape scale. So far, SiVM has been used successfully for water stress detection and prediction on the individual plant scale [26,61,62]. In this study, its applicability to the ecosystem scale was evaluated.

Before factorisation, all of the image cubes were transposed and concatenated along the band axis, which results in a data matrix $\mathbf{V}^{d\times n}$ with n samples of d-dimensional vectors. With SiVM, $\mathbf{V}^{d\times n}$ was decomposed into a basis (archetype) matrix $\mathbf{W}^{d\times k}$ and a coefficient matrix $\mathbf{H}^{k\times n}$ by iterative distance computations only. Thurau et al. [40] showed that fitting a $(k-1)$-simplex of maximal volume to the d-dimensional input data minimizes the Frobenius norm $\|\mathbf{V}-\mathbf{HW}\|$, a common optimisation target in matrix factorisation. The simplex approximates the convex hull of the data cloud, and it's vertices are the archetypical data points in $\mathbf{W}$. When compared to solving the quadratic optimisation problem as proposed by Cutler and Breiman [60] in their work on Archetypal Analysis, this approximation is computationally more efficient, and it minimizes the residual of $\mathbf{H}$ while preserving convexity: For $h_i = h_1, ..., h_k,\ h_{ij} \geq 0$ and $\sum_i h_{ij} = 1$. The number of archetypes

should be low enough to enable generalisation, but also high enough to cover a range of different extreme reflectance signatures. In this study, the number of archetypes was set to 30. To develop a qualitative classification of grassland stress levels from the latent components, the reflectance signatures in **W** were classified into three categories. Some archetypes resembled the spectra of healthy plants, others showed signs of stress, and a third group comprised spectra of non-vegetated pixels (background). The classification of archetypes was based on the results of field and lab experiments [63–68], visual assessment of archetype maps, and expert knowledge.

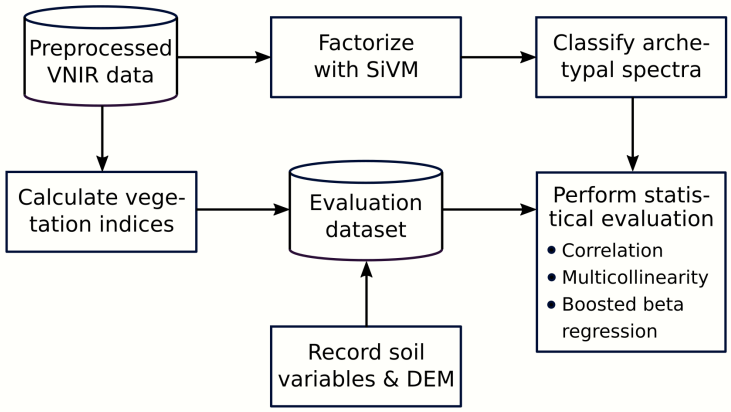


Figure 2. Overview of the main methodological workflow.

From a probabilistic viewpoint, the archetypes span a probability space in which each data point can be expressed as a convex combination of the archetypes. In other words, the coefficients h_{ij} in **H** provide a measure of similarity between any input pixel and the archetypical reflectance spectra in **W** [26]. Therefore, all of the data points in **X** can be viewed as draws from a specific Dirichlet distribution. This distribution is often used for proportional data and it has the advantage of imposing the convexity constraint on the coefficients in **H** [61]. The Dirichlet has one parameter α, which is a k-dimensional vector and it can be estimated with a maximum-likelihood approach as described by Minka [69]. The Dirichlet has a useful aggregation property that allows to merge parts of the sample space. For example, if one tosses a six-sided dice, then the probability of all rollable numbers can be described by $Dir(\alpha_1, ..., \alpha_6)$. Now, if the goal is to obtain the probability of rolling odd and even numbers, the aggregated (two-event) sample space still follows a Dirichlet distribution with aggregated parameter $Dir(\alpha_1 + \alpha_3 + \alpha_5, \alpha_2 + \alpha_4 + \alpha_6)$ [70]. In the same manner, archetypes were lumped together according to their category, "healthy", "stressed", or "background". This allowed creating maps of drought stress by summarizing coefficient values of healthy, stressed, and background archetypes, which are abbreviated as ζ, ν and ξ below. Each resulting aggregated archetype is still beta-distributed. Factorisation was performed with the pymf module (v0.3) [71].

2.5. Evaluation with Spectral Indices and Geophysical Measurements

Unsupervised learning techniques cluster datasets by exploring structures and patterns in the data and do not rely on ground truth (labels) for model building. Since vegetation measurements were not available for the research site DE-GsB, an internal validation of model performance was not feasible. An external evaluation that was based on a set of reliable proxy variables was implemented instead. This set consists of vegetation indices, which capture spectral plant traits. These traits are proxies for drought-inhibited plant growth resulting from changes to different biochemical, physiological, and structural plant traits [72]. A number of geophysical measurements as proxies for different soil traits and a high-resolution DEM derived from UAV data were included in the evaluation data set. We selected VNIR vegetation indices that either have been used in stress detection before or are related to ecophysiological traits impacted by drought stress, although those

relationships are not always linear. For most of them, successful application to grassland sites had been reported in the literature (see the following paragraphs). The indices range in complexity and use a variety of distinct spectral feature combinations. Table 2 provides details about the geophysical measurements and the DEM used in the evaluation dataset, while Table 3 gives an overview about the used vegetation indices.

Table 2. Overview of the variables in the evaluation dataset, which were not derived from the hyperspectral imagery.

Recording Date	Type of Measurement	Derived Variable
6 Mar. 2014	Electromagnetic induction with EM31	ECa (mS/m)
6 Mar. 2014	Electromagnetic induction with EM38	ECa (mS/m)
6 Mar. 2014	Gamma-ray spectrometry	Thorium^{232} (ppm)
6 Mar. 2014	Gamma-ray spectrometry	Dose rate (nGy/h)
18 Feb. 2015	Photogrammetry	DEM (m.a.s.l.)

The broadband indices Chl_{REopt} and Car_{REopt} are used as proxies for vegetation chlorophyll (C_{ab}) and carotenoid content (C_{xc}). The indices were developed by Féret et al. [24] using statistical models and model inversion on a large number of experimental and synthetic datasets in order to integrate variability between species. The red edge normalised difference vegetation index (RENDVI) was used to include an established chlorophyll-related vegetation index. The index is applicable for a broad range of species and it does not saturate for high-chlorophyll cases like dense canopies [73].

The PRI is correlated with short-term (minutes) changes of light use efficiency (LUE) through tracking variations in the de-epoxidation state of xantophylls, and water-stress indicators, like stomatal conductance [74–76]. It has proven to be applicable across a wide range of different types of leaf morphology, photosystems, and ecosystems. Beyond that, the PRI allows for the detection of long-term (weeks to months) changes of LUE caused by shifts in the C_{xc}/C_{ab} ratio [77–79] and is, therefore, correlated with seasonal variations in net CO_2 uptake [80]. Plant stress results in an increase of the C_{xc}/C_{ab} ratio and a decrease of photosynthetic LUE, which makes PRI well-suited for water stress detection. When compared to the original formulation, the PRI_{512} developed by Hernández-Clemente et al. [75] is less sensitive to structural effects. We assume the former index to have very similar proxy capabilities because the PRI_{512} has the same diagnostic band as the PRI.

The combination of a chlorophyll-sensitive red edge band and a NIR band makes the CTR2 an effective ratio index for stress detection [66,81]. The improved modified chlorophyll absorption ratio index (MCARI2) is a chlorophyll-corrected vegetation index for green LAI estimation [82]. In contrast to its predecessors, the index is less sensitive to variability in C_{ab} concentrations, and it has performed well when applied to grasslands for LAI estimation [83]. The improved modified soil-adjusted vegetation index (MSAVI2) [84] was included in the analysis to account for potential soil effects. It has been applied for LAI estimation and drought detection on different grassland sites [85,86]. Because MSAVI2 is a broadband index, it was calculated from the hyperspectral data based on Sentinel-2 band specifications.

Changes in RWC can be tracked with the WBI under progressive water stress [87,88]. However, studies found that derivative spectra of the slope of the NIR water absorption feature outperformed the reflectance-based water indices in RWC estimation in different experimental settings, including grassland sites in the Netherlands and UK [89,90]. WBI and the first-order derivative of the reflectance at 950.6 nm ($f'(\rho_{950.6})$) were both added to the set of evaluation variables. Derivatives were calculated by applying Savitzky–Golay smoothing with a window size of 67 nm and second-order polynomial fitting [91].

Table 3. List of vegetation indices used for evaluation of the SiVM-based stress level estimates (ρ denotes reflectance at a specific wavelength or wavelength range).

Index Name	Abbr.	Equation	Ref.
Opt. chlorophyll red edge index	$Chl_{RE\ opt}$	$(\rho_{680–730}{}^{-1} - \rho_{780–800}{}^{-1}) \times \rho_{755–780}$	[24]
Opt. carotenoid red edge index	$Car_{RE\ opt}$	$(\rho_{510–530}{}^{-1} - \rho_{680–730}{}^{-1}) \times \rho_{760–780}$	[24]
Red edge normalised difference vegetation index	RENDVI	$(\rho_{750} - \rho_{705})/(\rho_{750} + \rho_{705})$	[73]
Photochemical reflectance index 512	PRI_{512}	$(\rho_{531} - \rho_{512})/(\rho_{531} + \rho_{512})$	[75]
Carter index 2	CTR2	ρ_{695}/ρ_{760}	[81]
Modified chlorophyll absorption ratio index 2	MCARI2	$\frac{1.5[2.5(\rho_{800}-\rho_{670})-1.3(\rho_{800}-\rho_{550})]}{\sqrt{(2\rho_{800}+1)^2-(6\rho_{800}-5\sqrt{\rho_{670}})-0.5}}$	[82]
Modified soil-adjusted vegetation index 2	MSAVI2	$\frac{(2\rho_{NIR}+1-\sqrt{(2\rho_{NIR}+1)^2-8(\rho_{NIR}-\rho_{RED})})}{2}$	[84]
Water band index	WBI	ρ_{900}/ρ_{970}	[87]
First derivative @ 950.6 nm	-	$f'(\rho_{950.6})$	[90]

2.6. Statistical Inference

Spatial autocorrelation should be minimised for the statistical evaluation of the latent variable model. Therefore, a variogram analysis was conducted, followed by a systematic subsampling of data points. Background archetypes were used to exclude data points with $\xi \geq 0.4$. After the removal of defective (all-zero) pixels, the evaluation data set comprised 832 data points from both 2018 and 2019. Statistical analyses were carried out with R (v3.6.3) [92]. In a first step, a Pearson correlation matrix was computed to assess the agreement between ζ, ν and the evaluation data set using the package corrplot (v0.84) [93]. In addition, the matrix gave a first impression about the occurrence of multicollinearity among evaluation variables.

A regression approach was adopted for statistical inference and multicollinearity testing. Because ζ, ν are limited to the standard unit interval $(0, 1)$, beta regression, as introduced by Ferrari and Cribari-Neto [94], was used. Their parameterisation of the beta density uses μ as the mean of the response and ϕ as a precision parameter. The density of y is then defined as

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, 0 < y < 1, \quad (1)$$

where $0 < \mu < 1$, $\phi > 0$, and $\Gamma(\cdot)$ is the gamma function. For regressing the dependent variable on a set of independent variables, the logit link is used. The estimation of coefficients and distribution parameters is based on maximum likelihood. Beta regression was carried out using the betareg package (v3.1-3) [95]. Multicollinearity in the fitted model was assessed with the mctest package (v1.3.1) [96]. The condition number of the regressor matrix **X** and variance inflation factors (VIF) were used as criteria to determine collinearity among predictor variables. The objective of the statistical inference was to determine those variables (and, hence, plant traits) that most strongly influence the stress signal of the DE-GsB grassland site.

Multicollinearity leads to unstable coefficient estimates and biased statistics [97]. The detection of multicollinearity in the fitted beta regression model required an alternative model fitting approach that is less sensitive to collinearity. Boosted beta regression is a statistical learning technique that is suitable for this purpose. The algorithm is based on the principles of gradient boosting [98]: In an iterative process, weak learners are fit to the data to maximise the negative gradient of a given loss function. The model is updated in every iteration, eventually yielding a strong learner. The implementation of boosted beta regression uses the gamboostLSS [99] framework, which is an extension of generalised additive models for location, scale, and shape (GAMLSS), as introduced by Stasinopoulos and Rigby [100]. In contrast to many classical methods, like generalised additive models (GAM), this flexible, semiparametric approach to regression does not restrict the response variable to follow an exponential family distribution. Moreover, it is not restricted to modeling the conditional mean of the response. Instead, every distribution parameter can be modeled as a function of different predictor variables. In the case of

beta regression, this allows accounting for overdispersion by regressing ϕ on one or more predictor variables [101]. The predictor-response relationships are not limited to linear functions, but also encompass nonlinear and smooth functions, e.g., regression splines. The additive predictors are computed in the same way as in traditional GAMs [102].

Component-wise fitting of covariates constitutes the link between gradient boosting and GAMLSS. In the multivariate case, only one component of the regressor matrix $\mathbf{X}$ is fitted to the gradient vector per weak learner [103]. In each boosting iteration, the model is updated with the best-performing weak learner, which leads to a strictly additive structure [99]. For boosted beta regression, a weak learner will typically be a method, like univariate linear regression, penalised regression splines, or ridge regression. Because all of the distribution parameters are estimated, the algorithm constructs a distinct additive predictor for each distribution parameter by component-wise updates of separate prediction functions per iteration. Optimizing beta regression with gamboostLSS can be written as

$$(\hat{\mu}, \hat{\phi}) \leq \underset{\eta_\mu, \eta_\phi}{\arg\min} \left(\mathbb{E}_{\mathbf{Y},\mathbf{X}} \left[\rho \{ \mathbf{Y}, \eta_\mu(\mathbf{X}), \eta_\phi(\mathbf{X}) \} \right] \right), \tag{2}$$

where η_μ and η_ϕ are the additive predictors, ρ is the negative log-likelihood, and $(\mathbf{X}, \mathbf{Y})$ are predictor matrix and response vector [99]. Although, in practice, instead of minimizing the negative log-likelihood, the loss function is minimised by gradient descent, as the expected value is unknown. When setting an early stopping iteration m_{stop}, this approach has some beneficial properties: covariates that are never selected in the updating process are excluded from the final model [104]. Thus, data-driven variable selection to exclude less important predictors is performed, which alleviates collinearity [101]. Moreover, the coefficients of the selected predictors are shrunk towards zero. Similar to regularised regression approaches, early stopping in gamboostLSS introduces an estimation bias in exchange for decreased variance, which results in a more parsimonious model [105]. In low-dimensional settings with few predictors and many observations, boosting techniques are known to include too many noisy variables [106]. Therefore, m_{stop} was set to 300 to achieve sufficient variable selection and shrinkage, and the learning rate was set to 0.05. Three different models were fit: besides the full (interannual) model for the combined 2018–2019 dataset, two subset models were fit using only data points from 2018 or 2019. The evaluation datasets that were used in fitting of the models were mean-centered and standardised independently, and a global intercept variable was added. Boosted beta regression was performed using the gamboostLSS package (v2.0-1.1) [107].

3. Results

3.1. Drought Status

We investigated the drought and soil water status of the research site with three-months SPEI values and relative soil water content measurements (see Figure 3). The SPEI time series shows that the 2018 hyperspectral flight took place at the onset of an exceptional meteorological drought period that lasted until early 2019. At the time of first image acquisition, soil moisture had started to decrease. However, after a rather cool and wet summer and winter in 2017, soil moisture at 30 cm depth was still around 50%. During spring 2018, soil moisture was decreasing rapidly and SPEI values of ≤ -2 show persistent extreme drought conditions in summer 2018. The second image acquisition flight took place in spring 2019 after a winter with near-average precipitation and temperature conditions and subsequent re-emerging drought conditions, as indicated by the SPEI values. The soil moisture time series gives insight into the impact of the 2018–2019 drought on water availability at DE-GsB as the relative water content was quickly approaching the low levels of 2018 again. Because water reservoirs were insufficiently replenished by winter precipitation, the second image acquisition date is characterised by rather low soil water contents at around 25 or 35%. The analysis of these drought indicators confirms that the two hyperspectral scenes show contrasting environmental conditions that can be deemed to

be sufficiently extreme for application of SiVM. It also illustrates the intensity and duration of the drought period that affected the research site between the two flight campaigns.

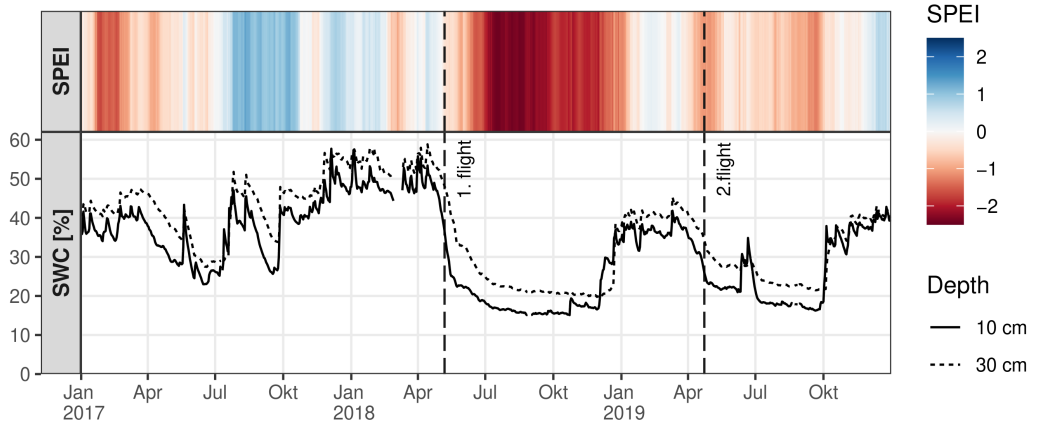


Figure 3. Time series of 91-days/3-months standardised precipitation-evapotranspiration index (SPEI) and relative soil water content (SWC) values for the research site DE-GsB. Vertical dashed lines indicate the dates of hyperspectral imaging flights.

3.2. Archetypes and Stress Maps

Matrix factorisation with SiVM resulted in a set of archetypes that encompass a range of spectra with distinct features. Figure 4 shows the classified archetypes. The background class holds more than half of all calculated archetypes due to the wide range of observed non-vegetation land cover. They mostly lack the trough in red reflectance and the sharp increase in the red edge area, which are characteristic for vegetation reflectance spectra [108]. Differences between healthy and moderately stressed vegetation are subtle. Vegetation stress leads to an increase in red reflectance, which reduces the red edge slope until the red reflectance trough around 680 nm disappears [23]. Healthy archetypes show a decrease in NIR reflectance beyond 900 nm, which is not the case for most stressed archetypes.

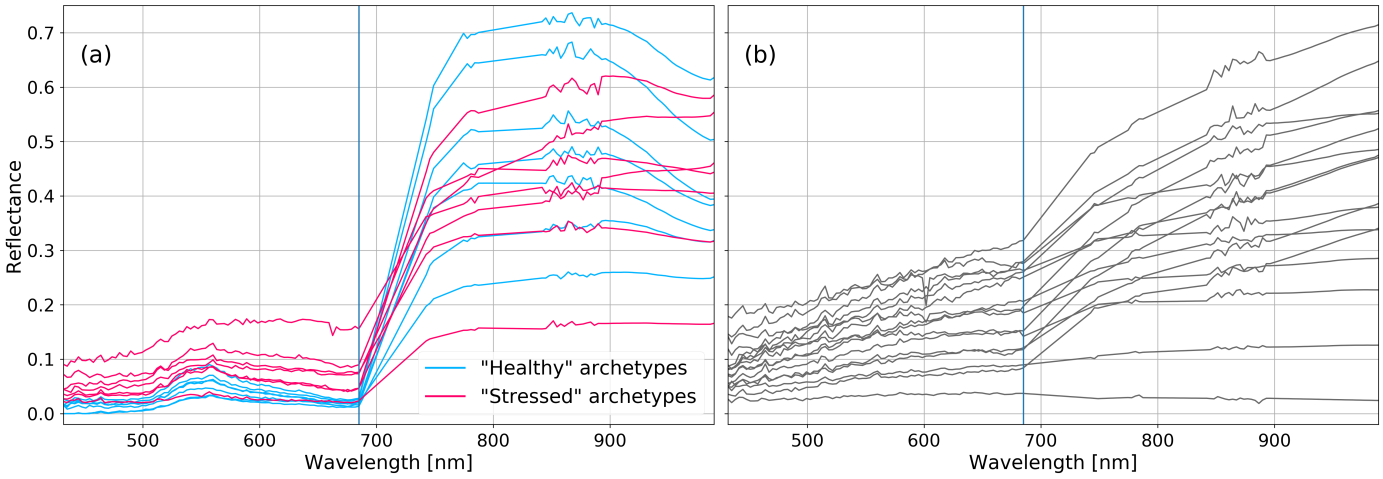
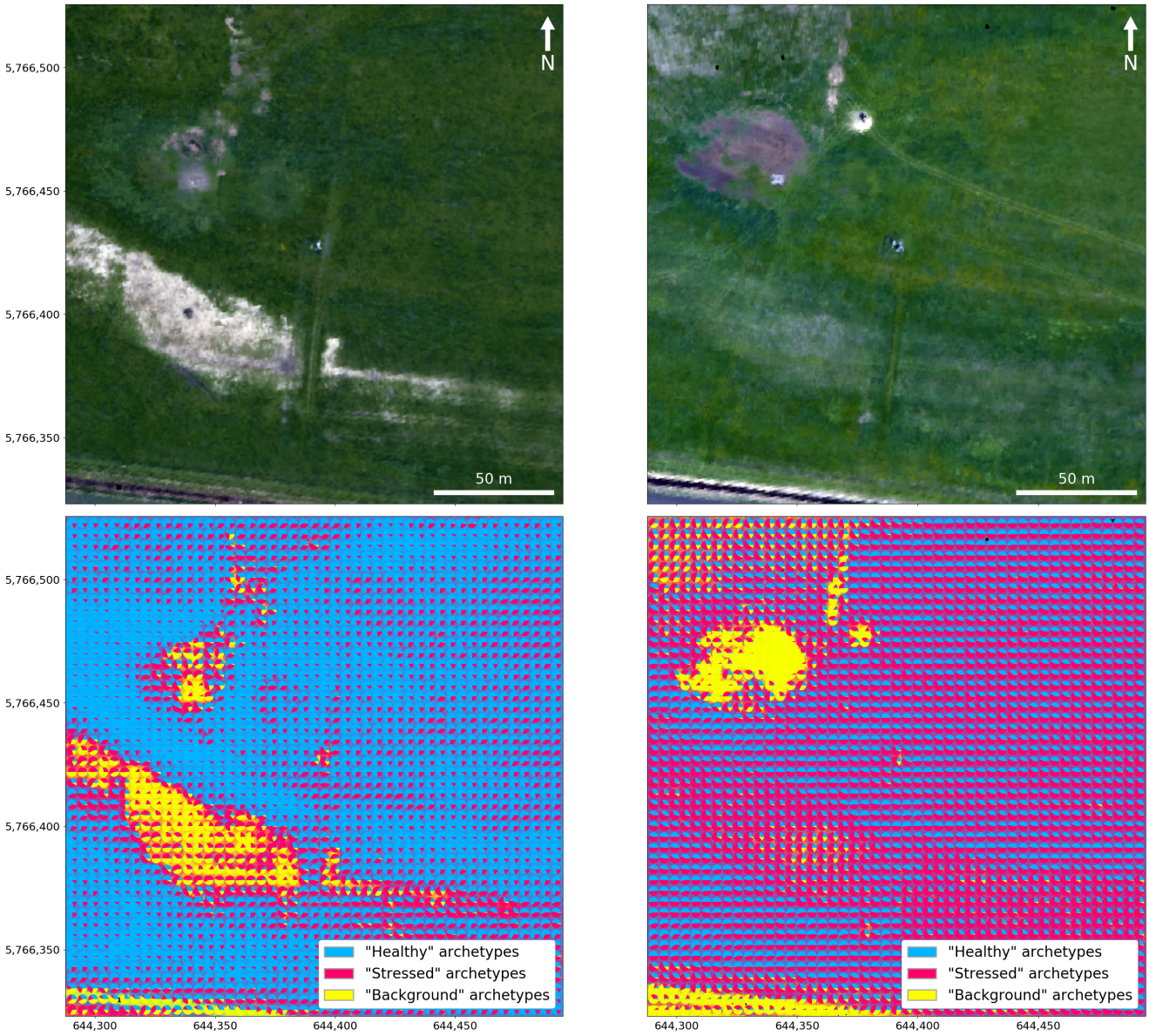


Figure 4. Classified archetypical reflectance spectra from the research site "Am Grossen Bruch" (DE-GsB). (**a**) Archetypes classified as "healthy" or "stressed". (**b**) All of the non-vegetation spectra were classified as "background". The vertical blue line marks the beginning of the red edge slope at 685 nm.

Figure 5 shows the distribution of archetype classes in hybrid maps using double layer visualisation [53] next to RGB representations of the same map section. For the image taken at the onset of the 2018–2019 drought, the average values for ζ (healthy) and ν (stressed) are 68.6 and 26.1%. The corresponding values for the image acquired one year later are 37.7

and 56.6%, which indicates intensified drought conditions. Both scenes contain patches of background signals also well visible in the true color images.



(**a**) 7 May 2018 (**b**) 23 April 2019

Figure 5. Maps of the research site "Am Grossen Bruch" (DE-GsB). RGB representations were computed with the HSI2RGB python module [109]. The double layer maps show the distribution of archetype classes. Archetype coefficient values per pixel always add up to one due to the convexity constraint of simplex volume maximisation (CRS: EPSG:32632).

3.3. Correlation Analysis

The results shown in Figure 6 show an overall good agreement of the computed qualitative stress metric and vegetation indices, implying a robust performance of SiVM-based stress classification at the canopy level. The correlation matrix of the aggregated archetypes ζ and ν, as well as the evaluation variables, demonstrates the correspondence between the latent variable-based stress index and established vegetation growth proxies as well as potentially relevant geophysical measurements. Correlation analysis of the spectral indices revealed considerable collinearity among the predictor variables. The absolute correlations of ζ are consistently stronger than for ν. We assume this to be a consequence of the stronger similarity between stressed and background spectra that makes an exact

distinction of these categories more difficult. Therefore, we focused on ζ as a response variable in the statistical analyses. The RENDVI as a proxy for C_{ab} content and plant vitality is positively correlated with ζ. Likewise, the two optimised pigment indices Chl_{REopt} and Car_{REopt} point to higher pigment concentrations in less stressed vegetation, with Car_{REopt} exhibiting the strongest correlation of all evaluation variables ($r = 0.89, p < 0.001$). Of the two vegetation indices that are related to structural traits, the MCARI2 exhibited slightly higher correlation values than the MSAVI2. The results indicate a larger photosynthetically active leaf area in vegetation with higher ζ values, as expected. The RWC-related spectral indices WBI and $f'(\rho_{950.6})$ lead to very similar results and suggest a good level of agreement between the latent variable-based stress index and plant water content.

The correspondence between the ECa and GR measurements, as well as stress level estimates ζ/ν, is low. For these soil proxy variables, the strongest correlations result from the EM38 measurements ($r = -0.28, p < 0.001$). The DEM exhibits similarly low correlation values. Geophysical variables and DEM seem to be rather unrelated to the computed stress signal. Nevertheless, the variables were included in the statistical model to verify this assertion by variable selection.

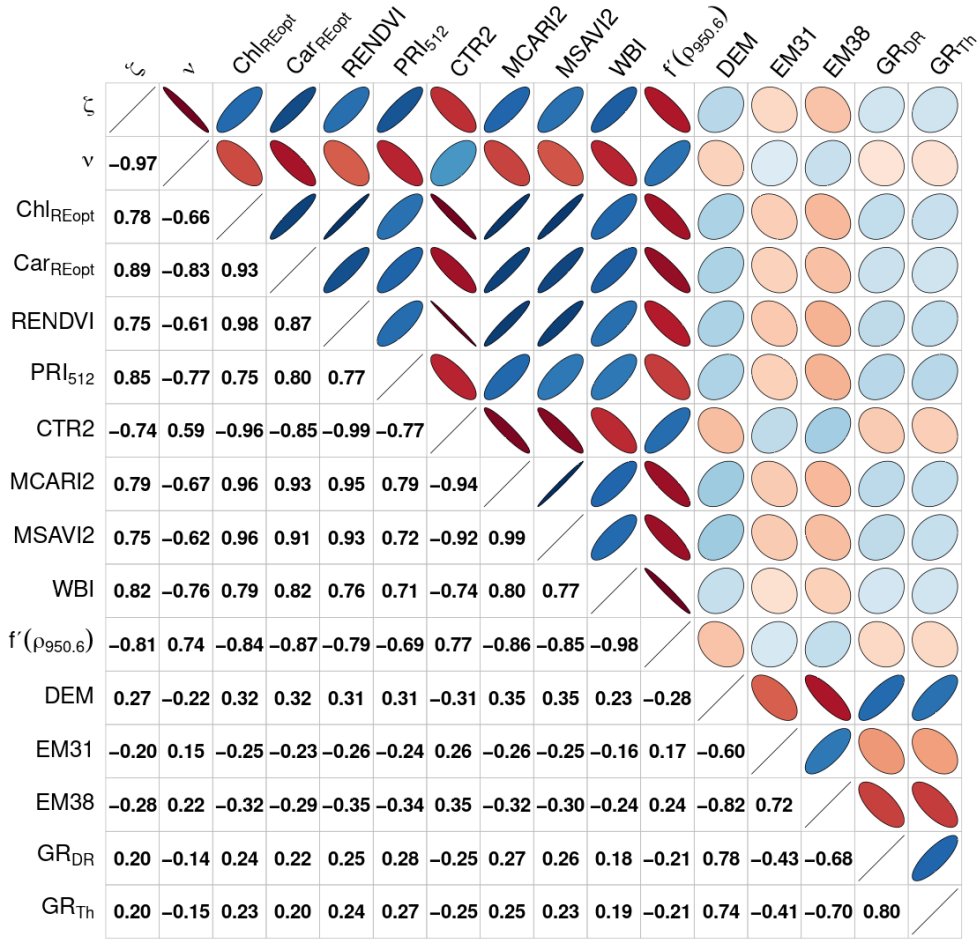


Figure 6. The correlation matrix of stress metrics ζ, ν and the evaluation dataset, which includes spectral indices, geophysical measurements and a digital elevation model (DEM). All of the correlations are highly significant ($p < 0.001$).

3.4. Boosted Beta Regression

Before interpreting the results of the statistical analysis, multicollinearity diagnostics were performed with a beta regression model that included ζ as the dependent and all evaluation variables as the independent variables. The beta distribution parameters μ and ϕ were both estimated by the full set of regressors. The regressor matrix has a condition

number of 3388, which suggests severe multicollinearity. VIF also indicate multicollinearity in all vegetation indices, except for the PRI_{512}, which had a VIF of 6.8 (see Table 4). While Car_{REopt} had a VIF of 19, all other vegetation index VIF values were in the range of 73–216, which indicates severe multicollinearity. The rest of the predictor variables exhibited substantially lower VIF values.

Table 4 shows the estimated coefficient values from the three different model fits for the expected value μ and the precision parameter ϕ. The interannual full model (FM) results show that substantial variable selection has taken place. Four out of nine spectral indices that were related to plant traits were not included in the estimation of μ. MCARI2, CTR2, and RENDVI were excluded or marginalised in all μ models. Likewise, all of the geophysical variables and DEM were either excluded or shrunk towards zero in all μ models. Car_{REopt} is the most important predictor with the highest μ coefficient value in all models. Thus, increased C_{xc} is associated with an increased probability of observing a healthy pixel. The same applies to higher PRI_{512} values in FM μ. This dependency is much weaker in the subset models M18 μ and M19 μ. Increased WBI values are associated with less vegetation stress in all models, with weaker relationships in the subset models. WBI and $f'(\rho_{950.6})$ had been almost perfectly collinear ($r = -0.98$) in the correlation analysis, so that the effect of the derivative reflectance did not contribute to FM μ. The situation is less clear in the subset models where both of the predictors that are related to relative water content contribute to the regression, although $f'(\rho_{950.6})$ has relatively small coefficient values. In contrast to the results of the correlation analysis, the coefficients for Chl_{REopt} and MSAVI2 are negative in M18 μ and M19 μ. e.g., MSAVI2 is related to green LAI, but a correspondence of lower LAI and healthier vegetation is implausible. The inclusion of collinear variables induces instability in the model and it can result in erratic coefficient estimation. We assume this to be the case for Chl_{REopt} and MSAVI2, as both are highly correlated with Car_{REopt}. In the subset models, MSAVI2 shows small, negative coefficient values, while the picture is not clear for Chl_{REopt}. In conclusion, the strongest predictors in the full model are linked to carotenoids (Car_{REopt}, PRI_{512}), LAI (MSAVI2), and water content (WBI). The subset μ models contain three main predictors: Car_{REopt}, Chl_{REopt}, and WBI. The difference is especially large for the PRI_{512} which is only relevant in the FM μ. Coefficients for M19 μ are generally smaller, probably due to the prevalence of stressed vegetation and, hence, the smaller range of stress values in the 2019 dataset.

Table 4. Variance inflation factors (VIF) from multicollinearity analysis of the evaluation variables. The estimated coefficient values from fitting the interannual full model (FM) and subset models (M18, M19) with boosted beta regression (early stopping after 300 iterations, learning rate: 0.05, coefficient values rounded to three digits).

Name	VIF	FM μ	M18 μ	M19 μ	FM ϕ	M18 ϕ	M19 ϕ
Chl_{REopt}	88.4	−0.143	0.163	−0.215	−0.071	0	−0.390
Car_{REopt}	19.1	1.018	0.721	0.435	−1.005	−1.587	0
RENDVI	214.3	0	0.052	0	0	0	0
PRI_{512}	6.8	0.322	0.011	0.079	0.009	0.495	0
CTR2	139.0	0	−0.031	0	−0.949	−1.065	−0.406
MCARI2	147.4	0	0	0	0	0.160	0
MSAVI2	138.8	−0.396	−0.035	−0.077	0.154	0	0.091
WBI	73.2	0.265	0.244	0.143	0	0.004	0.179
$f'(\rho_{950.6})$	104.3	0	0.136	−0.038	0	0	0
DEM	5.1	0.004	−0.023	0.010	−0.048	−0.160	0.025
EM31	2.2	0	0.042	−0.002	−0.063	−0.200	0
EM38	5.2	0	−0.032	−0.071	−0.251	−0.204	−0.149
GR_{DR}	3.6	0	0.004	0	0	0	0.076
GR_{Th}	3.4	−0.009	−0.043	−0.014	0	0.012	0
Intercept		0.014	0.206	0	2.356	2.566	1.324

Car_{REopt} is the most relevant predictor in two out of three ϕ models. ϕ is a precision parameter and, therefore, vegetation with higher Car_{REopt} values is associated with a larger variance of y. In contrast to the μ models, CTR2 was an important predictor in

every ϕ model. Likewise, the DEM and geophysical variables have small, mostly negative coefficients in the precision models, with the EM38 measurements being the most relevant. However, in this study, focus was on the location parameter μ.

4. Discussion

4.1. Archetype Classification

The classification of archetypical spectra shown in Figure 4 is in line with results from previous experiments. Several studies reported decreases in NIR reflectance beyond 850 nm under increasing water stress [63,64,66]. Likewise, the water absorption feature around 970 nm becomes less pronounced [110]. The decrease in NIR reflectance over 900 nm exhibited by healthy archetypes is probably related to stronger water absorption in healthy vegetation. Nevertheless, NIR reflectance exhibits much more variability with varying biophysical variables, like leaf angle distribution or LAI, when compared to the visible spectrum [111]. Therefore, distinguishing healthy and stressed spectra is more reliable when based on visible spectrum wavelengths [108]. Further criteria like NIR reflectance should be consulted in the case of ambiguity because a classification of mixed pixels cannot be highly selective.

The maps presented in Figure 5 show strongly increased vegetation stress levels across the entire site in the 2019 image. This can be attributed to the prolonged drought in the region (see Figure 3). The substantial decrease in soil water availability has a marked impact on the pasture. The background patch that is shown in the upper left quarter of both images is related to grazing and management of the site, while the extended background patch in the lower left quarter of the 2018 image can be explained by a depression that is regularly flooded. In Figure 5b, vegetation in the depression exhibits among the highest stress levels owing to a combination of multiple stress agents. This shows that the computed stress index does not distinguish between the two stressors that occurred in the area of investigation before the second flight campaign.

The analysis of field data entails limited control of influencing factors. In the case of the DE-GsB grassland site, this is particularly the case for grazing. Depending on its intensity, grazing can be an additional stressor, but it can also have beneficial effects, such as a reduction of competition between plants [112]. Because detailed information about grazing intensity and timing was not available, we did not include the influence of grazing in the analysis. Most grass species are shallow-rooted, so that their water status is more dependent on surface soil water and, hence, precipitation [113]. Therefore, long-term stress has different consequences for grassland species, which are often adapted to cope with periods of water shortage when compared to deep-rooted plants. For such a dynamic ecosystem, the comparison of the 2018 and 2019 stress classification should, thus, not be interpreted as the aggregated effect of a prolonged drought period. The datasets rather represent two extreme cases that can be viewed as reference points to observe anomalies.

4.2. Importance of Variables

We validated the latent variable-based stress classification with a set of evaluation variables. As a general remark, these results are not based on direct vegetation measurements, which limits the informative value of the computed relationships. Nevertheless, the results of the correlation analysis imply the successful application of SiVM at the ecosystem scale as the inferred stress signal has a high degree of correspondence with changes in plant trait proxies indicating inhibited growth. Best-performing vegetation indices are Car_{REopt} (proxy for C_{xc}), PRI_{512} (proxy for C_{xc}/C_{ab} ratio and LUE) and WBI (proxy for RWC). For assessing the severity of vegetation stress, pigment-related indices Chl_{REopt}, Car_{REopt}, and PRI_{512} are important. Carotenoids serve multiple purposes in plant physiology, playing an important role in capturing light energy as well as dissipating excess energy [114]. Because they protect the plants' photosystems from damage, e.g., counteracting different forms of reactive oxygen species, increases in C_{xc} are usually observed under severe stress conditions [115]. Therefore, we expected an increase in Car_{REopt} (and hence C_{xc}) in stressed

vegetation, but the lower Car_{REopt} values in the 2019 dataset imply only moderate stress conditions. However, one would expect C_{ab} to decrease faster than total C_{xc} [73,116] under stress conditions. Such an increase in the C_{xc}/C_{ab} ratio would manifest in a tendency towards lower PRI values [78]. In fact, the strong positive correlation of ζ and PRI_{512} suggests an increase in the C_{xc}/C_{ab} ratio following the drought period.

These results can be explained by the sensitivity of grasslands to precipitation patterns [117], which allows for faster recovery after rainfall events. Although soil water was decreasing again in spring 2019, winter precipitation had allowed some replenishment of the diminishing reservoirs (see Figure 3). The timing of data collection is another influencing factor. Both of the images were taken in spring, so that the ecosystems were not subject to the heat stress of the summer months. We conclude that the average vegetation stress level at DE-GsB in spring 2019 was not severe, but moderate. This claim is also supported by the shape of the archetypes in Figure 4. Severe stress leads to a notable "blue shift" of the red edge transition point that moves to shorter wavelengths [65,118]. No such shift could be observed in the stressed archetypes with the given spectral resolution of 3.2 nm.

Boosted beta regression was useful in removing redundant predictors from the model, despite some presumed remaining instability in the final models due to multicollinearity [119,120]. Nevertheless, statistical inference supports the conclusions regarding the variable importance drawn from the correlation analysis. All of the geophysical measurements and the DEM were excluded from all μ models or shrunk towards zero. In case of the DEM, the exclusion can be attributed to the level research site which only has an elevation range of about 1 m within it's 200 $\times$ 200 m extent. The estimation of the contribution of soil traits to the vegetation stress signal would require the integration of actual soil traits, such as soil moisture, porosity or pH. Keeping the strong collinearity among certain vegetation indices in mind, the notion of "irrelevant" proxy variables should be considered with caution [121]. If there is a physiological cause for collinearity among predictors, the variables in question have to be considered as relevant. This is the case for the pigment indices Chl_{REopt} and Car_{REopt} (see Figure 6). We assume a close correlation of C_{ab} and C_{xc} as long as plants are not severely stressed [122], even though Car_{REopt} is the vegetation index that shows the closest agreement with ζ values. PRI_{512} is the predictor variable with the smallest VIF value in multicollinearity testing and it is an important variable in the full μ model. However, it was quite irrelevant in the subset models. This demonstrates that the PRI_{512} tracked "long-term" stress responses of the grassland site, but it is less related to the within-image stress gradient. We reason that the PRI_{512} contributed most unique information to the estimation procedure. The importance of this spectral trait constitutes a possible link between the latent variable model and photosynthesis.

4.3. Potential of SiVM for Vegetation Monitoring

Even though there is a strong correlation between ζ and the carotenoid-related variables Car_{REopt} and PRI_{512}, additional campaigns to collect imagery of severely stressed grassland would be required to confirm the results. This illustrates a restriction of SiVM-based stress detection that does not result in a standardised, quantitative measurement of stress, but generates a qualitative classification of stress levels occuring in the input datasets. Hyperspectral reference datasets of extreme environmental conditions can be included in an analysis to address this issue.

The present work also highlights several advantages of the factorisation-based stress detection approach. The integrated treatment of background, shade, and other effects that can adversely affect an analysis makes different preprocessing steps obsolete. Because SiVM is a method of unsupervised learning, it does not require previous knowledge or labeled data to perform clustering. This property makes it a valuable tool for areas where ground truth is not available, location of spectral features of interest is unknown, or in the case of rapid diagnosis of stress levels of vegetation, irrespective of the actual driver. Given that all calculated archetypes and their combinations are restricted to the unit interval, results from

the factorisation are easily and intuitively interpretable. Although not investigated in this study, SiVM has proven to be successful in pre-symptomatic stress detection, being able to track temporal stress development several days earlier when compared to visual analysis or vegetation indices [26]. The method has also been expanded to include Gaussian process priors of arbitrary covariates, such as plant species or location for prediction tasks [61].

The demonstrated approach to latent variable stress detection did not result in a distinction of different stressors. The fact that drought and waterlogging both lead to an increased production of reactive oxygen species and changes in carotenoid content [123] is one example for the difficulties in stress agent distinction. The reflectance changes in the visible wavelengths are known to be similar for different stress sources [65]. Stress responses are species dependent, but the 0.4 m spatial resolution in this analysis did not allow for detecting species responses. Thus, we inferred the functional stress response of a grassland ecosystem. The quality and specificity of stress classifications could be improved by incorporating SWIR and TIR wavelengths into the factorisation [26]. Both of the spectral regions contain essential information regarding the vegetation health status related to heat and drought stress, and they have been employed in many applications [22,124,125]. Nevertheless, the visible spectrum is a reliable indicator of plant stress in general. The use of visible spectrum sensors is cheaper and, therefore, more widespread, which makes the methodology presented in this study suitable for low-cost monitoring networks [126], whether in precision agriculture or scientific applications [127].

The generalisation of the methodology has to be further tested. As a first step, SiVM should be applied to imagery from different ecosystems to evaluate the robustness of the resulting stress index and investigate necessary adjustments of the approach. While the method has performed satisfactory using data from a rather homogeneous canopy of simple-structured grass leaves, its application to a structurally complex canopy remains to be evaluated. Extrapolation of derived stress indices to larger scales has to be evaluated in detail. Different applications of SiVM for stress detection at the leaf scale have already been proven successful [26,61,128]. In this study, we could show that the methodology can be transferred to the ecosystem scale. Hence, an upscaling effort to the regional or continental scale based on spaceborne data seems to be worthwhile, when considering the relevance of satellites for monitoring tasks and the increasing ability of next-gen spaceborne spectrometers, like DESIS [129], PRISMA [130], or the upcoming EnMAP [131].

5. Conclusions

The main achievement of this study is the successful application of SiVM for an unsupervised classification of grassland drought stress at the ecosystem scale. The stress-related vegetation indices included in the analysis were strongly correlated with the stress classification computed with SiVM (four out of nine with $|r| > 0.8$, none with $|r| < 0.75$, $p < 0.001$). Previously, this method has been applied at a close range above individual plants and a homogeneous corn canopy [26,61]. The findings of this study suggest that SiVM is robust towards mixed pixel effects, and it can be transferred to field scale applications and beyond. This property, the capability for pre-symptomatic stress detection, and the easily interpretable results make SiVM a valuable tool for precision agriculture applications that use remote sensing data from airborne sensors. Large-scale stress detection is also highly valuable for reducing uncertainties in land surface and ecosystem models. Physiological stress indices can be integrated to improve the estimation of the impact of stress on carbon and water fluxes of ecosystems [132,133].

We performed different statistical analyses to assess the quality of SiVM-based stress classification using a set of evaluation variables that included spectral indices and geophysical measurements. In the correlation analysis, Car_{REopt} had the highest correlation with computed vegetation stress levels ($r = 0.89$, $p < 0.001$). This proxy for carotenoid content indicates higher carotenoid levels in healthier vegetation. In contrast to our expectations, Car_{REopt} values in the 2019 dataset decreased, which suggested a C_{xc} reduction. This shows that neither prolonged water stress during 2018 nor the re-emering drought in spring 2019

had a lasting or severe impact on vegetation health at DE-GsB. Immediately after an intensive heat and drought event, like the summer 2018 drought, increased C_{xc} values would most likely have been observable. However, a validation with direct vegetation measurements would have generated more reliable results.

Statistical inference with boosted beta regression revealed that PRI_{512}, the second carotenoid-related index, is closely related to the unsupervised stress classification in the interannual full model and it shows comparably low multicollinearity with other evaluation variables. Likewise, covariation between PRI and carbon uptake of plants has been found in a number of studies [74,80,134]. These results suggest that a relationship between the SiVM-based stress classification and photosynthetic efficiency exists. Thus, the methodology could possibly be used for the estimation of carbon fluxes based on hyperspectral data. Further research on the issue would require local hyperspectral data of high temporal resolution in combination with eddy covariance measurements to enable the calibration of remotely sensed flux estimates [25]. Local scale optical sensors are less dependent on favorable atmospheric conditions than satellites and enable continuous measurements. If carbon flux estimation with SiVM proves to be practical at the local scale, transfer to satellite scale could be considered for large-scale productivity monitoring.

Plant stress that is caused by water scarcity or water logging cannot be distinguished so far. This limitation could be improved by introducing more categories into the archetype classification. However, a more precise distinction requires in-depth knowledge regarding the subtle disparities of reflectance spectra due to different stress agents for different species and landscape types. A database of labeled spectral fingerprints of stress effects could be used to achieve a more differentiated classification with high accuracy using supervised learning techniques. Beyond that, labeling data within already operative "fingerprinting" efforts, like the Spectranomics database [135], would produce synergies that could foster an easier usage of hyperspectral data for different monitoring tasks.

Author Contributions: Conceptualisation, F.H., U.W. and A.L.; methodology, F.H. and F.P.; software, F.H.; validation, F.H.; formal analysis, F.H. and F.P.; investigation, F.H.; resources, C.R.; data curation, F.H., C.R. and G.S.; writing—original draft preparation, F.H.; writing—review and editing, F.H., F.P., C.R., G.S., U.W. and A.L.; visualisation, F.H. and F.P.; supervision, U.W. and A.L.; project administration, C.R. and A.L.; funding acquisition, C.R., U.W. and A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available in the UFZ data archive at http://www.ufz.de/record/dmp/archive/10591 (accessed on 25 March 2021, CC BY-NC-SA 4.0).

Acknowledgments: This work was supported by the Helmholtz Association in the framework of MOSES (Modular Observation Solutions for Earth Systems, https://www.ufz.de/moses/, accessed on 27 March 2021). Conceptualization and realisation of this study were conducted within the MOMENT project (Model Monitoring EveNTs, https://www.ufz.de/index.php?en=46352, accessed on 27 March 2021). The authors are thankful for the good cooperation and inspiring discussions within the project group. We kindly thank F. Boeing for providing interpolated daily meteorological data and S. Kögler for providing the DEM. We thank M. Pohle for acquisition of geophysical data. The technical support from A. Schmidt and M. Wu is gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

C_{ab}	Chlorophyll content
C_{xc}	Carotenoid content
$Car_{RE\ opt}$	Optimised carotenoid red edge index
$Chl_{RE\ opt}$	Optimised chlorophyll red edge index
CRS	Coordinate reference system
CTR2	Carter index 2
DE-GsB	TERENO/ICOS site "Am Grossen Bruch"
DEM	Digital elevation model
DESIS	DLR Earth Sensing Imaging System Spectrometers
ECa	Apparent electrical conductivity
EMI	Electromagnetic induction
EnMAP	Environmental Mapping and Analysis Program
FM	Full model
GAM	Generalised additive model
GAMLSS	Generalised additive models for location, scale and shape
GR	Gamma ray
ICOS	Integrated Carbon Observation System Research Infrastructure
LAI	Leaf area index
LUE	Light use efficiency
MCARI2	Modified chlorophyll absorption ratio index 2
MSAVI2	Modified soil-adjusted vegetation index 2
NDVI	Normalised difference vegetation index
NIR	Near-infrared
PRI	Photochemical reflectance index
PRISMA	PRecursore IperSpettrale della Missione Applicativa
RENDVI	Red edge normalised difference vegetation index
RWC	Relative water content
SiVM	Simplex volume maximisation
SPEI	Standardised precipitation-evapotranspiration index
SWC	Soil water content
SWIR	Short-wavelength infrared
TDR	Time-domain reflectometer
TERENO	TERrestrial ENvironmental Observatories
TIR	Thermal infrared
UAV	Unmanned aerial vehicle
VIF	Variance inflation factor
VNIR	Visible and near-infrared
WBI	Water band index

References

1. Field, C.B.; Barros, V.; Stocker, T.F.; Qin, D.; Dokken, D.J.; Ebi, K.L.; Mastrandrea, M.D.; Mach, K.J.; Plattner, G.K.; Allen, S.K.; et al. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*; A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change; Cambridge University Press: Cambridge, UK, 2012.
2. Spinoni, J.; Vogt, J.V.; Naumann, G.; Barbosa, P.; Dosio, A. Will Drought Events Become More Frequent and Severe in Europe? *Int. J. Climatol.* **2018**, *38*, 1718–1736. [CrossRef]
3. Herring, S.C.; Christidis, N.; Hoell, A.; Hoerling, M.P.; Stott, P.A. Explaining Extreme Events of 2018 from a Climate Perspective. *Bull. Am. Meteorol. Soc.* **2020**, *101*, S1–S140. [CrossRef]
4. Ionita, M.; Tallaksen, L.; Kingston, D.; Stagge, J.; Laaha, G.; Van Lanen, H.; Scholz, P.; Chelcea, S.; Haslinger, K. The European 2015 Drought from a Climatological Perspective. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 1397–1419. [CrossRef]
5. Hari, V.; Rakovec, O.; Markonis, Y.; Hanel, M.; Kumar, R. Increased Future Occurrences of the Exceptional 2018–2019 Central European Drought under Global Warming. *Sci. Rep.* **2020**, *10*, 1–10. [CrossRef]
6. Zscheischler, J.; Mahecha, M.D.; Von Buttlar, J.; Harmeling, S.; Jung, M.; Rammig, A.; Randerson, J.T.; Schölkopf, B.; Seneviratne, S.I.; Tomelleri, E. A Few Extreme Events Dominate Global Interannual Variability in Gross Primary Production. *Environ. Res. Lett.* **2014**, *9*, 035001. [CrossRef]

7. Lowe, A.; Harrison, N.; French, A.P. Hyperspectral Image Analysis Techniques for the Detection and Classification of the Early Onset of Plant Disease and Stress. *Plant Methods* **2017**, *13*, 80. [CrossRef]
8. Guimarães, N.; Pádua, L.; Marques, P.; Silva, N.; Peres, E.; Sousa, J.J. Forestry Remote Sensing from Unmanned Aerial Vehicles: A Review Focusing on the Data, Processing and Potentialities. *Remote Sens.* **2020**, *12*, 1046. [CrossRef]
9. Feld, C.K.; Sousa, J.P.; Da Silva, P.M.; Dawson, T.P. Indicators for Biodiversity and Ecosystem Services: Towards an Improved Framework for Ecosystems Assessment. *Biodivers. Conserv.* **2010**, *19*, 2895–2919. [CrossRef]
10. Lausch, A.; Bastian, O.; Klotz, S.; Leitão, P.J.; Jung, A.; Rocchini, D.; Schaepman, M.E.; Skidmore, A.K.; Tischendorf, L.; Knapp, S. Understanding and Assessing Vegetation Health by in Situ Species and Remote-Sensing Approaches. *Methods Ecol. Evol.* **2018**, *9*, 1799–1809. [CrossRef]
11. Schrodt, F.; Bailey, J.J.; Kissling, W.D.; Rijsdijk, K.F.; Seijmonsbergen, A.C.; van Ree, D.; Hjort, J.; Lawley, R.S.; Williams, C.N.; Anderson, M.G.; et al. Opinion: To Advance Sustainable Stewardship, We Must Document Not Only Biodiversity but Geodiversity. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 16155–16158. [CrossRef] [PubMed]
12. Zhu, Z.; Wulder, M.A.; Roy, D.P.; Woodcock, C.E.; Hansen, M.C.; Radeloff, V.C.; Healey, S.P.; Schaaf, C.; Hostert, P.; Strobl, P.; et al. Benefits of the Free and Open Landsat Data Policy. *Remote Sens. Environ.* **2019**, *224*, 382–385. [CrossRef]
13. Vereecken, H.; Weihermüller, L.; Jonard, F.; Montzka, C. Characterization of Crop Canopies and Water Stress Related Phenomena Using Microwave Remote Sensing Methods: A Review. *Vadose Zone J.* **2012**, *11*, vzj2011.0138ra. [CrossRef]
14. Asner, G.P.; Brodrick, P.G.; Anderson, C.B.; Vaughn, N.; Knapp, D.E.; Martin, R.E. Progressive Forest Canopy Water Loss during the 2012–2015 California Drought. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E249–E255. [CrossRef]
15. Chen, J.M.; Liu, J.; Leblanc, S.G.; Lacaze, R.; Roujean, J.L. Multi-Angular Optical Remote Sensing for Assessing Vegetation Structure and Carbon Absorption. *Remote Sens. Environ.* **2003**, *84*, 516–525. [CrossRef]
16. Vilfan, N.; Van der Tol, C.; Muller, O.; Rascher, U.; Verhoef, W. Fluspect-B: A Model for Leaf Fluorescence, Reflectance and Transmittance Spectra. *Remote Sens. Environ.* **2016**, *186*, 596–615. [CrossRef]
17. Schimel, D.; Pavlick, R.; Fisher, J.B.; Asner, G.P.; Saatchi, S.; Townsend, P.; Miller, C.; Frankenberg, C.; Hibbard, K.; Cox, P. Observing Terrestrial Ecosystems and the Carbon Cycle from Space. *Glob. Chang. Biol.* **2015**, *21*, 1762–1776. [CrossRef]
18. Martin, R.E.; Chadwick, K.D.; Brodrick, P.G.; Carranza-Jimenez, L.; Vaughn, N.R.; Asner, G.P. An Approach for Foliar Trait Retrieval from Airborne Imaging Spectroscopy of Tropical Forests. *Remote Sens.* **2018**, *10*, 199. [CrossRef]
19. Zhu, X.; Cai, F.; Tian, J.; Williams, T.K.A. Spatiotemporal Fusion of Multisource Remote Sensing Data: Literature Survey, Taxonomy, Principles, Applications, and Future Directions. *Remote Sens.* **2018**, *10*, 527. [CrossRef]
20. Govender, M.; Govender, P.J.; Weiersbye, I.M.; Witkowski, E.T.F.; Ahmed, F. Review of Commonly Used Remote Sensing and Ground-Based Technologies to Measure Plant Water Stress. *Water SA* **2009**, *35*, 741–752. [CrossRef]
21. Barton, C. Advances in Remote Sensing of Plant Stress. *Plant Soil* **2012**, *354*, 41–44. [CrossRef]
22. Hernández-Clemente, R.; Hornero, A.; Mottus, M.; Penuelas, J.; González-Dugo, V.; Jiménez, J.C.; Suárez, L.; Alonso, L.; Zarco-Tejada, P.J. Early Diagnosis of Vegetation Health from High-Resolution Hyperspectral and Thermal Imagery: Lessons Learned from Empirical Relationships and Radiative Transfer Modelling. *Curr. For. Rep.* **2019**, *5*, 169–183. [CrossRef]
23. Jones, H.G.; Vaughan, R.A. *Remote Sensing of Vegetation: Principles, Techniques, and Applications*; Oxford University Press: New York, NY, USA, 2010.
24. Féret, J.B.; François, C.; Gitelson, A.; Asner, G.P.; Barry, K.M.; Panigada, C.; Richardson, A.D.; Jacquemoud, S. Optimizing Spectral Indices and Chemometric Analysis of Leaf Chemical Properties Using Radiative Transfer Modeling. *Remote Sens. Environ.* **2011**, *115*, 2742–2750. [CrossRef]
25. Garbulsky, M.F.; Peñuelas, J.; Gamon, J.; Inoue, Y.; Filella, I. The Photochemical Reflectance Index (PRI) and the Remote Sensing of Leaf, Canopy and Ecosystem Radiation Use Efficiencies: A Review and Meta-Analysis. *Remote Sens. Environ.* **2011**, *115*, 281–297. [CrossRef]
26. Römer, C.; Wahabzada, M.; Ballvora, A.; Pinto, F.; Rossini, M.; Panigada, C.; Behmann, J.; Léon, J.; Thurau, C.; Bauckhage, C.; et al. Early Drought Stress Detection in Cereals: Simplex Volume Maximisation for Hyperspectral Image Analysis. *Funct. Plant Biol.* **2012**, *39*, 878–890. [CrossRef] [PubMed]
27. Peerbhay, K.Y.; Mutanga, O.; Ismail, R. Random Forests Unsupervised Classification: The Detection and Mapping ofSolanum mauritianumInfestations in Plantation Forestry Using Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3107–3122. [CrossRef]
28. Behmann, J.; Steinrücken, J.; Plümer, L. Detection of Early Plant Stress Responses in Hyperspectral Images. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 98–111. [CrossRef]
29. Ceamanos, X.; Valero, S. Processing Hyperspectral Images. In *Optical Remote Sensing of Land Surface*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2016; pp. 163–200.
30. Lee, D.D.; Seung, H.S. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature* **1999**, *401*, 788–791. [CrossRef]
31. Pauca, V.P.; Shahnaz, F.; Berry, M.W.; Plemmons, R.J. Text Mining Using Non-Negative Matrix Factorizations. In Proceedings of the 2004 SIAM International Conference on Data Mining, Lake Buena Vista, FL, USA, 22–24 April 2004; pp. 452–456.

[CrossRef]
32. Li, Y.; Ngom, A. The Non-Negative Matrix Factorization Toolbox for Biological Data Mining. *Source Code Biol. Med.* **2013**, *8*, 10. [CrossRef] [PubMed]
33. Lee, J.H.; Hashimoto, R.; Wible, C.G.; Yoo, S.S. Investigation of Spectrally Coherent Resting-State Networks Using Non-Negative Matrix Factorization for Functional MRI Data. *Int. J. Imaging Syst. Technol.* **2011**, *21*, 211–222. [CrossRef]
34. Ball, N.M.; Brunner, R.J. Data Mining and Machine Learning in Astronomy. *Int. J. Mod. Phys. D* **2010**, *19*, 1049–1106. [CrossRef]
35. Jia, S.; Qian, Y. Constrained Nonnegative Matrix Factorization for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2008**, *47*, 161–173. [CrossRef]
36. Gillis, N.; Plemmons, R.J. Dimensionality Reduction, Classification, and Spectral Mixture Analysis Using Non-Negative Underapproximation. *Opt. Eng.* **2011**, *50*, 027001. [CrossRef]
37. Karoui, M.S.; Deville, Y.; Hosseini, S.; Ouamri, A.; Ducrot, D. Contribution of Non-Negative Matrix Factorization to the Classification of Remote Sensing Images. In Proceedings of the Image and Signal Processing for Remote Sensing XIV. International Society for Optics and Photonics, Cardiff, Wales, UK, 15–18 September 2008; Volume 7109, p. 71090X. [CrossRef]
38. Huang, B.; Song, H.; Cui, H.; Peng, J.; Xu, Z. Spatial and Spectral Image Fusion Using Sparse Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 1693–1704. [CrossRef]
39. Danaher, S.; O'mongain, E. Singular Value Decomposition in Multispectral Radiometry. *Int. J. Remote Sens.* **1992**, *13*, 1771–1777. [CrossRef]
40. Thurau, C.; Kersting, K.; Bauckhage, C. Yes We Can: Simplex Volume Maximization for Descriptive Web-Scale Matrix Factorization. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM'10, Toronto, ON, Canada, 25–29 October 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 1785–1788. [CrossRef]
41. Vicente-Serrano, S.M.; Beguería, S.; López-Moreno, J.I. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *J. Clim.* **2010**, *23*, 1696–1718. [CrossRef]
42. Wollschläger, U.; Attinger, S.; Borchardt, D.; Brauns, M.; Cuntz, M.; Dietrich, P.; Fleckenstein, J.H.; Friese, K.; Friesen, J.; Harpke, A. The Bode Hydrological Observatory: A Platform for Integrated, Interdisciplinary Hydro-Ecological Research within the TERENO Harz/Central German Lowland Observatory. *Environ. Earth Sci.* **2017**, *76*, 29. [CrossRef]
43. Zacharias, S.; Bogena, H.; Samaniego, L.; Mauder, M.; Fuß, R.; Pütz, T.; Frenzel, M.; Schwank, M.; Baessler, C.; Butterbach-Bahl, K.; et al. A Network of Terrestrial Environmental Observatories in Germany. *Vadose Zone J.* **2011**, *10*, 955–973. [CrossRef]
44. Rebmann, C.; Aubinet, M.; Schmid, H.; Arriga, N.; Aurela, M.; Burba, G.; Clement, R.; De Ligne, A.; Fratini, G.; Gielen, B. ICOS Eddy Covariance Flux-Station Site Setup: A Review. *Int. Agrophys.* **2018**, *32*, 471–494. [CrossRef]
45. Bernhofer, C.; Goldberg, V.; Franke, J.; Surke, M.; Adam, J. *Regionale Klimadiagnose für Sachsen-Anhalt, Abschlussbericht zum Forschungsvorhaben des Landesamtes für Umweltschutz Sachsen-Anhalt*; Berichte des Landesamtes für Umweltschutz Sachsen-Anhalt; Technische Universität Dresden: Dresden, Germany, 2008; Volume 5.
46. Beguería, S.; Vicente-Serrano, S.M.; Reig, F.; Latorre, B. Standardized Precipitation Evapotranspiration Index (SPEI) Revisited: Parameter Fitting, Evapotranspiration Models, Tools, Datasets and Drought Monitoring. *Int. J. Climatol.* **2014**, *34*, 3001–3023. [CrossRef]
47. Vergni, L.; Todisco, F.; Mannocchi, F. Evaluating the Uncertainty and Reliability of Standardized Indices. *Hydrol. Res.* **2017**, *48*, 701–713. [CrossRef]
48. Zink, M.; Kumar, R.; Cuntz, M.; Samaniego, L. A High-Resolution Dataset of Water Fluxes and States for Germany Accounting for Parametric Uncertainty. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 1769–1790. [CrossRef]
49. Kumar, R.; Musuuza, J.L.; Loon, A.F.V.; Teuling, A.J.; Barthel, R.; Ten Broek, J.; Mai, J.; Samaniego, L.; Attinger, S. Multiscale Evaluation of the Standardized Precipitation Index as a Groundwater Drought Indicator. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 1117–1131. [CrossRef]
50. Schläpfer, D.; Richter, R. Geo-Atmospheric Processing of Airborne Imaging Spectrometry Data. Part 1: Parametric Orthorectification. *Int. J. Remote Sens.* **2002**, *23*, 2609–2630. [CrossRef]
51. Richter, R.; Schläpfer, D. Geo-Atmospheric Processing of Airborne Imaging Spectrometry Data. Part 2: Atmospheric/Topographic Correction. *Int. J. Remote Sens.* **2002**, *23*, 2631–2649. [CrossRef]
52. DWD Climate Data Center. *Historical Hourly Weather Station Measurements of Visibility in Germany*; Version v002; Deutscher Wetterdienst: Offenbach, Germany, 2019.
53. Cai, S.; Du, Q.; Moorhead, R.J. Hyperspectral Imagery Visualization Using Double Layers. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3028–3036. [CrossRef]
54. Cook, P.G.; Hughes, M.W.; Walker, G.R.; Allison, G.B. The Calibration of Frequency-Domain Electromagnetic Induction Meters and Their Possible Use in Recharge Studies. *J. Hydrol.* **1989**, *107*, 251–265. [CrossRef]
55. Lausch, A.; Zacharias, S.; Dierke, C.; Pause, M.; Kühn, I.; Doktor, D.; Dietrich, P.; Werban, U. Analysis of Vegetation and Soil Patterns Using Hyperspectral Remote Sensing, EMI, and Gamma-Ray Measurements. *Vadose Zone J.* **2013**, *12*, 1–15. [CrossRef]
56. Brogi, C.; Huisman, J.A.; Pätzold, S.; von Hebel, C.; Weihermüller, L.; Kaufmann, M.S.; van der Kruk, J.; Vereecken, H. Large-Scale Soil Mapping Using Multi-Configuration EMI and Supervised Image Classification. *Geoderma* **2019**, *335*, 133–148. [CrossRef]
57. Martini, E.; Werban, U.; Zacharias, S.; Pohle, M.; Dietrich, P.; Wollschläger, U. Repeated Electromagnetic Induction Measurements for Mapping Soil Moisture at the Field Scale: Validation with Data from a Wireless Soil Moisture Monitoring Network. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 495. [CrossRef]

58. Müller, S.; Schüler, L.; Zech, A.; Attinger, S.; Heße, F. GeoStat-Framework/GSTools: V1.2.1. *Zenodo* **2020**. [CrossRef]
59. Ding, C.H.; Li, T.; Jordan, M.I. Convex and Semi-Nonnegative Matrix Factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *32*, 45–55. [CrossRef] [PubMed]
60. Cutler, A.; Breiman, L. Archetypal Analysis. *Technometrics* **1994**, *36*, 338–347. [CrossRef]
61. Kersting, K.; Xu, Z.; Wahabzada, M.; Bauckhage, C.; Thurau, C.; Roemer, C.; Ballvora, A.; Rascher, U.; Leon, J.; Pluemer, L. Pre-Symptomatic Prediction of Plant Drought Stress Using Dirichlet-Aggregation Regression on Hyperspectral Images. In Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI-12), Toronto, ON, Canada, 22–26 July 2012; AAAI Press: Toronto, ON, Canada, 2012; Volume 26, pp. 302–308. [CrossRef]
62. Kersting, K.; Wahabzada, M.; Römer, C.; Thurau, C.; Ballvora, A.; Rascher, U.; Léon, J.; Bauckhage, C.; Plümer, L. Simplex Distributions for Embedding Data Matrices over Time. In Proceedings of the 2012 SIAM International Conference on Data Mining, Anaheim, CA, USA, 26–28 April 2012; pp. 295–306. [CrossRef]
63. Kim, Y.; Glenn, D.M.; Park, J.; Ngugi, H.K.; Lehman, B.L. Hyperspectral Image Analysis for Plant Stress Detection. In Proceedings of the 2010 ASABE Annual International Meeting, Pittsburgh, PA, USA, 20–23 June 2010; American Society of Agricultural and Biological Engineers: Pittsburgh, PA, USA, 2010; p. 1. [CrossRef]
64. El-Hendawy, S.; Al-Suhaibani, N.; Hassan, W.; Tahir, M.; Schmidhalter, U. Hyperspectral Reflectance Sensing to Assess the Growth and Photosynthetic Properties of Wheat Cultivars Exposed to Different Irrigation Rates in an Irrigated Arid Region. *PLoS ONE* **2017**, *12*, e0183262. [CrossRef] [PubMed]
65. Carter, G.A. Responses of Leaf Spectral Reflectance to Plant Stress. *Am. J. Bot.* **1993**, *80*, 239–243. [CrossRef]
66. Bayat, B.; Van der Tol, C.; Verhoef, W. Remote Sensing of Grass Response to Drought Stress Using Spectroscopic Techniques and Canopy Reflectance Model Inversion. *Remote Sens.* **2016**, *8*, 557. [CrossRef]
67. Aldakheel, Y.Y.; Danson, F.M. Spectral Reflectance of Dehydrating Leaves: Measurements and Modelling. *Int. J. Remote Sens.* **1997**, *18*, 3683–3690. [CrossRef]
68. Vogelmann, J.E.; Rock, B.N.; Moss, D.M. Red Edge Spectral Measurements from Sugar Maple Leaves. *Int. J. Remote Sens.* **1993**, *14*, 1563–1575. [CrossRef]
69. Minka, T. *Estimating a Dirichlet Distribution*; Technical Report; MIT: Cambridge, MA, USA, 2000.
70. Frigyik, B.A.; Kapila, A.; Gupta, M.R. *Introduction to the Dirichlet Distribution and Related Processes*; Technical Report; University of Washington: Washington, DC, USA, 2010.
71. Thurau, C. Python Matrix Factorization Module. 2014. Available online: https://github.com/cthurau/pymf (accessed on 11 May 2021).
72. Lausch, A.; Erasmi, S.; King, D.J.; Magdon, P.; Heurich, M. Understanding Forest Health with Remote Sensing, Part I: A Review of Spectral Traits, Processes and Remote-Sensing Characteristics. *Remote Sens.* **2016**, *8*, 1029. [CrossRef]
73. Gitelson, A.; Merzlyak, M.N. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus Hippocastanum L. and Acer Platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *J. Plant Physiol.* **1994**, *143*, 286–292. [CrossRef]
74. Gamon, J.; Serrano, L.; Surfus, J.S. The Photochemical Reflectance Index: An Optical Indicator of Photosynthetic Radiation Use Efficiency across Species, Functional Types, and Nutrient Levels. *Oecologia* **1997**, *112*, 492–501. [CrossRef]
75. Hernández-Clemente, R.; Navarro-Cerrillo, R.M.; Suárez, L.; Morales, F.; Zarco-Tejada, P.J. Assessing Structural Effects on PRI for Stress Detection in Conifer Forests. *Remote Sens. Environ.* **2011**, *115*, 2360–2375. [CrossRef]
76. Zarco-Tejada, P.J.; González-Dugo, V.; Williams, L.E.; Suárez, L.; Berni, J.A.; Goldhamer, D.; Fereres, E. A PRI-Based Water Stress Index Combining Structural and Chlorophyll Effects: Assessment Using Diurnal Narrow-Band Airborne Imagery and the CWSI Thermal Index. *Remote Sens. Environ.* **2013**, *138*, 38–50. [CrossRef]
77. Ustin, S.L.; Roberts, D.A.; Gamon, J.A.; Asner, G.P.; Green, R.O. Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. *BioScience* **2004**, *54*, 523–534. [CrossRef]
78. Filella, I.; Porcar-Castell, A.; Munné-Bosch, S.; Bäck, J.; Garbulsky, M.F.; Peñuelas, J. PRI Assessment of Long-Term Changes in Carotenoids/Chlorophyll Ratio and Short-Term Changes in de-Epoxidation State of the Xanthophyll Cycle. *Int. J. Remote Sens.* **2009**, *30*, 4443–4455. [CrossRef]
79. Garrity, S.R.; Eitel, J.U.; Vierling, L.A. Disentangling the Relationships between Plant Pigments and the Photochemical Reflectance Index Reveals a New Approach for Remote Estimation of Carotenoid Content. *Remote Sens. Environ.* **2011**, *115*, 628–635. [CrossRef]
80. Stylinski, C.; Gamon, J.; Oechel, W. Seasonal Patterns of Reflectance Indices, Carotenoid Pigments and Photosynthesis of Evergreen Chaparral Species. *Oecologia* **2002**, *131*, 366–374. [CrossRef]
81. Carter, G.A. Ratios of Leaf Reflectances in Narrow Wavebands as Indicators of Plant Stress. *Remote Sens.* **1994**, *15*, 697–703. [CrossRef]
82. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral Vegetation Indices and Novel Algorithms for Predicting Green LAI of Crop Canopies: Modeling and Validation in the Context of Precision Agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352. [CrossRef]
83. Li, Z.; Guo, X. A Suitable Vegetation Index for Quantifying Temporal Variation of Leaf Area Index (LAI) in Semiarid Mixed Grassland. *Can. J. Remote Sens.* **2010**, *36*, 709–721. [CrossRef]

84. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]
85. Liu, Z.Y.; Huang, J.F.; Wu, X.H.; Dong, Y.P. Comparison of Vegetation Indices and Red-Edge Parameters for Estimating Grassland Cover from Canopy Reflectance Data. *J. Integr. Plant Biol.* **2007**, *49*, 299–306. [CrossRef]
86. Wu, Z.; Velasco, M.; McVay, J.; Middleton, B.; Vogel, J.; Dye, D. MODIS Derived Vegetation Index for Drought Detection on the San Carlos Apache Reservation. *Int. J. Adv. Remote Sens. GIS* **2016**, *5*, 1524–1538. [CrossRef]
87. Penuelas, J.; Filella, I.; Serrano, L.; Save, R. Cell Wall Elasticity and Water Index (R970 Nm/R900 Nm) in Wheat under Different Nitrogen Availabilities. *Int. J. Remote Sens.* **1996**, *17*, 373–382. [CrossRef]
88. Peñuelas, J.; Inoue, Y. Reflectance Indices Indicative of Changes in Water and Pigment Contents of Peanut and Wheat Leaves. *Photosynthetica* **1999**, *36*, 355–360. [CrossRef]
89. Rollin, E.M.; Milton, E.J. Processing of High Spectral Resolution Reflectance Data for the Retrieval of Canopy Water Content Information. *Remote Sens. Environ.* **1998**, *65*, 86–92. [CrossRef]
90. Clevers, J.G.; Kooistra, L.; Schaepman, M.E. Using Spectral Information from the NIR Water Absorption Features for the Retrieval of Canopy Water Content. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 388–397. [CrossRef]
91. Savitzky, A.; Golay, M.J. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
92. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
93. Wei, T.; Simko, V. R Package "Corrplot": Visualization of a Correlation Matrix. 2017. Available online: https://CRAN.R-project.org/package=corrplot (accessed on 11 May 2021).
94. Ferrari, S.; Cribari-Neto, F. Beta Regression for Modelling Rates and Proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [CrossRef]
95. Cribari-Neto, F.; Zeileis, A. Beta Regression in R. *J. Stat. Softw.* **2010**, *34*, 1–24. [CrossRef]
96. Imdadullah, M.; Aslam, M.; Altaf, S. Mctest: An R Package for Detection of Collinearity among Regressors. *R. J.* **2016**, *8*, 495–505. [CrossRef]
97. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitao, P.J. Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Ecography* **2013**, *36*, 27–46. [CrossRef]
98. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
99. Mayr, A.; Fenske, N.; Hofner, B.; Kneib, T.; Schmid, M. Generalized Additive Models for Location, Scale and Shape for High Dimensional Data—a Flexible Approach Based on Boosting. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2012**, *61*, 403–427. [CrossRef]
100. Stasinopoulos, M.; Rigby, B. Generalized Additive Models for Location Scale and Shape. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2005**, *54*, 507–554. [CrossRef]
101. Schmid, M.; Wickler, F.; Maloney, K.O.; Mitchell, R.; Fenske, N.; Mayr, A. Boosted Beta Regression. *PLoS ONE* **2013**, *8*, e61623. [CrossRef]
102. Hastie, T.; Tibshirani, R. *Generalized Additive Models*; CRC Press: Boca Raton, FL, USA, 1990.
103. Bühlmann, P.; Yu, B. Boosting with the L 2 Loss: Regression and Classification. *J. Am. Stat. Assoc.* **2003**, *98*, 324–339. [CrossRef]
104. Mayr, A.; Hofner, B.; Schmid, M. The Importance of Knowing When to Stop. *Methods Inf. Med.* **2012**, *51*, 178–186. [CrossRef]
105. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer Science & Business Media: New York, NY, USA, 2009.
106. Bühlmann, P.; Gertheiss, J.; Hieke, S.; Kneib, T.; Ma, S.; Schumacher, M.; Tutz, G.; Wang, C.Y.; Wang, Z.; Ziegler, A. Discussion of "the Evolution of Boosting Algorithms" and "Extending Statistical Boosting". *Methods Inf. Med.* **2014**, *53*, 436–445. [CrossRef] [PubMed]
107. Hofner, B.; Mayr, A.; Schmid, M. gamboostLSS: An R Package for Model Building and Variable Selection in the GAMLSS Framework. *arXiv* **2014**, arXiv:1407.1774.
108. Knipling, E.B. Physical and Physiological Basis for the Reflectance of Visible and Near-Infrared Radiation from Vegetation. *Remote Sens. Environ.* **1970**, *1*, 155–159. [CrossRef]
109. Magnusson, M.; Sigurdsson, J.; Armannsson, S.; Ulfarsson, M.; Deborah, H.; Sveinsson, J. Creating RGB Images from Hyperspectral Images Using a Color Matching Function. In Proceedings of the 2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Waikoloa, HI, USA, 26 September–2 October 2020; pp. 2045–2048. [CrossRef]
110. Caturegli, L.; Matteoli, S.; Gaetani, M.; Grossi, N.; Magni, S.; Minelli, A.; Corsini, G.; Remorini, D.; Volterrani, M. Effects of Water Stress on Spectral Reflectance of Bermudagrass. *Sci. Rep.* **2020**, *10*, 15055. [CrossRef]
111. Asner, G.P. Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sens. Environ.* **1998**, *64*, 234–253. [CrossRef]
112. Bilotta, G.S.; Brazier, R.E.; Haygarth, P.M. The Impacts of Grazing Animals on the Quality of Soils, Vegetation, and Surface Waters in Intensively Managed Grasslands. *Adv. Agron.* **2007**, *94*, 237–280. [CrossRef]
113. Nippert, J.B.; Holdo, R.M. Challenging the Maximum Rooting Depth Paradigm in Grasslands and Savannas. *Funct. Ecol.* **2015**, *29*, 739–745. [CrossRef]
114. Hallik, L.; Kazantsev, T.; Kuusk, A.; Galmés, J.; Tomás, M.; Niinemets, Ü. Generality of Relationships between Leaf Pigment Contents and Spectral Vegetation Indices in Mallorca (Spain). *Reg. Environ. Chang.* **2017**, *17*, 2097–2109. [CrossRef]

115. Sudrajat, D.J.; Siregar, I.Z.; Khumaida, N.; Siregar, U.J.; Mansur, I. Adaptability of White Jabon (Anthocephalus Cadamba MIQ.) Seedling from 12 Populations to Drought and Waterlogging. *AGRIVITA J. Agric. Sci.* **2015**, *37*, 130–143. [CrossRef]
116. Penuelas, J.; Baret, F.; Filella, I. Semi-Empirical Indices to Assess Carotenoids/Chlorophyll a Ratio from Leaf Spectral Reflectance. *Photosynthetica* **1995**, *31*, 221–230.
117. Knapp, A.K.; Carroll, C.J.W.; Denton, E.M.; La Pierre, K.J.; Collins, S.L.; Smith, M.D. Differential Sensitivity to Regional-Scale Drought in Six Central US Grasslands. *Oecologia* **2015**, *177*, 949–957. [CrossRef] [PubMed]
118. Chen, J.; Li, F.; Wang, R.; Fan, Y.; Raza, M.A.; Liu, Q.; Wang, Z.; Cheng, Y.; Wu, X.; Yang, F.; et al. Estimation of Nitrogen and Carbon Content from Soybean Leaf Reflectance Spectra Using Wavelet Analysis under Shade Stress. *Comput. Electron. Agric.* **2019**, *156*, 482–489. [CrossRef]
119. Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735–743. [CrossRef]
120. Tomaschek, F.; Hendrix, P.; Baayen, R.H. Strategies for Addressing Collinearity in Multivariate Linguistic Data. *J. Phon.* **2018**, *71*, 249–267. [CrossRef]
121. Tuv, E. Feature Selection with Ensembles, Artificial Variables, and Redundancy Elimination. *J. Mach. Learn. Res.* **2009**, *10*, 1341–1366.
122. Gitelson, A.A.; Keydan, G.P.; Merzlyak, M.N. Three-Band Model for Noninvasive Estimation of Chlorophyll, Carotenoids, and Anthocyanin Contents in Higher Plant Leaves. *Geophys. Res. Lett.* **2006**, *33*. [CrossRef]
123. Uarrota, V.G.; Stefen, D.L.V.; Leolato, L.S.; Gindri, D.M.; Nerling, D. Revisiting Carotenoids and Their Role in Plant Stress Responses: From Biosynthesis to Plant Signaling Mechanisms during Stress. In *Antioxidants and Antioxidant Enzymes in Higher Plants*; Springer: Cham, Switzerland, 2018; pp. 207–232.
124. Zhang, N.; Hong, Y.; Qin, Q.; Liu, L. VSDI: A Visible and Shortwave Infrared Drought Index for Monitoring Soil and Vegetation Moisture Based on Optical Remote Sensing. *Int. J. Remote Sens.* **2013**, *34*, 4585–4609. [CrossRef]
125. Raza, S.e.A.; Smith, H.K.; Clarkson, G.J.J.; Taylor, G.; Thompson, A.J.; Clarkson, J.; Rajpoot, N.M. Automatic Detection of Regions in Spinach Canopies Responding to Soil Moisture Deficit Using Combined Visible and Thermal Imagery. *PLoS ONE* **2014**, *9*, e97612. [CrossRef] [PubMed]
126. Lausch, A.; Borg, E.; Bumberger, J.; Dietrich, P.; Heurich, M.; Huth, A.; Jung, A.; Klenke, R.; Knapp, S.; Mollenhauer, H. Understanding Forest Health with Remote Sensing, Part III: Requirements for a Scalable Multi-Source Forest Health Monitoring Network Based on Data Science Approaches. *Remote Sens.* **2018**, *10*, 1120. [CrossRef]
127. Jawad, H.M.; Nordin, R.; Gharghan, S.K.; Jawad, A.M.; Ismail, M. Energy-Efficient Wireless Sensor Networks for Precision Agriculture: A Review. *Sensors* **2017**, *17*, 1781. [CrossRef]
128. Kuska, M.; Wahabzada, M.; Leucker, M.; Dehne, H.W.; Kersting, K.; Oerke, E.C.; Steiner, U.; Mahlein, A.K. Hyperspectral Phenotyping on the Microscopic Scale: Towards Automated Characterization of Plant-Pathogen Interactions. *Plant Methods* **2015**, *11*, 28. [CrossRef]
129. Alonso, K.; Bachmann, M.; Burch, K.; Carmona, E.; Cerra, D.; De los Reyes, R.; Dietrich, D.; Heiden, U.; Hölderlin, A.; Ickes, J. Data Products, Quality and Validation of the DLR Earth Sensing Imaging Spectrometer (DESIS). *Sensors* **2019**, *19*, 4471. [CrossRef]
130. Loizzo, R.; Guarini, R.; Longo, F.; Scopa, T.; Formaro, R.; Facchinetti, C.; Varacalli, G. PRISMA: The Italian Hyperspectral Mission. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 175–178. [CrossRef]
131. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sens.* **2015**, *7*, 8830–8857. [CrossRef]
132. Pugh, T.A.; Rademacher, T.; Shafer, S.L.; Steinkamp, J.; Barichivich, J.; Beckage, B.; Haverd, V.; Harper, A.; Heinke, J.; Nishina, K. Understanding the Uncertainty in Global Forest Carbon Turnover. *Biogeosciences* **2020**, *17*, 3961–3989. [CrossRef]
133. Fisher, R.; McDowell, N.; Purves, D.; Moorcroft, P.; Sitch, S.; Cox, P.; Huntingford, C.; Meir, P.; Woodward, F.I. Assessing Uncertainties in a Second-Generation Dynamic Vegetation Model Caused by Ecological Scale Limitations. *New Phytol.* **2010**, *187*, 666–681. [CrossRef] [PubMed]
134. Peñuelas, J.; Filella, I.; Gamon, J.A. Assessment of Photosynthetic Radiation-Use Efficiency with Spectral Reflectance. *New Phytol.* **1995**, *131*, 291–296. [CrossRef]
135. Asner, G.P.; Martin, R.E. Spectranomics: Emerging Science and Conservation Opportunities at the Interface of Biodiversity and Remote Sensing. *Glob. Ecol. Conserv.* **2016**, *8*, 212–219. [CrossRef]